Hybridization and Application of Machine Comprehension Algorithms

A brief outline of the proposed research to be carried out

in pursuit of the degree of

Doctor of Philosophy

in

Computer Science

Area of Research

Soft Computing, Machine Comprehension

Submitted by

Bingi Binathi

under the supervision of

Dr. Lotika Singh

Faculty of Science

Dayalbagh Educational Institute

(Deemed University)

Dayalbagh, Agra – 282110

(Session: September 2017)

1. Introduction

Natural Language Processing (NLP) techniques combined with Artificial Intelligence (AI) algorithms have enabled machines to translate and paraphrase, generate questions and answer questions. These conversation systems developed for different text processing and analysis tasks can be brought under a broad domain of Machine Comprehension (MC). Machine Comprehension systems have vivid applications in the fields of education, commerce, human resourcing, etc. The thrust areas of research in the domain of MC are translation and paraphrasing, question generation and question answering.

- Translation and Paraphrasing: The two major translation models are Statistical Machine Translation (SMT) [1] and Rule-Based Machine Translation (RBMT) [2]. The paraphrasing operation includes reordering, lexical substitutions and syntactic transformations [1]. One of the well established techniques to build paraphrases is to leverage on bilingual corpora to find equivalent phrases in a single language by pivoting over a shared translation in another language [3].

- Question Generation (QG): Existing systems generate natural-language questions conditioned on answers. These conversational products use natural language processing (NLP) and neural network (NN) techniques to generate answer based questions but have a limited scope in determining and asking context based questions [4-7].

- Question Answering (QA): It is the task of answering a question given a corresponding document. It not only provides humans with efficient access to vast amounts of information but also acts as an important proxy task to assess machine literacy via reading comprehension [8].

1.1 Motivation

There is a significant progress in question answering but more work needs to be done in the areas of paraphrasing and question generation. For instance, in question generation domain, algorithms are able to generate and answer objective questions but are unable to frame context based subjective questions. Similarly, in machine comprehension, machines are able to read passages and paraphrase them using bilingual parallel corpora [9] but are far beyond the goal of paraphrasing using monolingual corpus [10].

2.0 Review of Paraphrasing Algorithms

Paraphrasing is generating different expressions with the same meaning as of the original content and is fundamental to many NLP tasks. Paraphrases allow creation of more fluent and vivid text [11]. Identification of paraphrases helps in removing redundancy in multi-document summarization [12]. Paraphrases help in addressing the issues faced during machine translation by providing an alternative and valid way to translate the text [13]. Identification of paraphrased text also help in answering questions and providing evidence that the answer is correct [10].

2.1 Statistical techniques

Paraphrase generation and identification algorithms had focused on the automatic extraction of paraphrases from different types of corpora consisting of parallel, non-parallel, and comparable texts. Monolingual corpora had been used for paraphrasing [13-15] but the usefulness was limited due to narrow range of availability. So, the models used bilingual parallel corpora to induce paraphrases based on techniques from phrase-based statistical machine translation [16]. Bannard and Callison-Burch's [17] developed bilingual pivoting algorithm. This method inputs two English strings e1 and e2 that translate to the same foreign string (f) which can be assumed to have the same meaning. The method then pivots over f to extract <e1, e2> as a pair of paraphrases [18]. Syntax based SMT technique ([9], [17]) was extended to

create syntactic paraphrases leading to the creation of paraphrase database (PPDB) [19-20], a large-scale paraphrase database containing over a billion of paraphrase pairs in 23 different languages.

Models that are independent of syntactic knowledge that could perform sentiment analysis were developed [21]. An immediate extension of such models could be parsing, parts-of-speech tagging and name entity recognition. Initial models were designed using statistical techniques and made use of monolingual corpus but could not succeed due to limited availability of data. These models inspired researchers to design phrase based - bilingual corpora models which contributed to better paraphrasing but couldn't resolve the semantic issues.

2.2 Soft Computing techniques

Neural machine translation (NMT) model could address the semantic issue as well as achieve state-of-the art performance using minimal linguistic information. NMT models based on neural networks were developed for semantic similarity identification between phrases [4-7]. NMT methods can also be ported with bilingual pivoting method and it offers three advantages over conventional methods [18]:

- Such models learn continuous space representations for phrases and sentences that can be incorporated in recognizing textual similarity and entailment tasks.
- It will either score a pair of potential paraphrase candidates or generate target paraphrases for a given input.
- The architecture of model enables generation to take advantage of wider context in comparison to phrase-based approaches.

Data-driven paraphrasing algorithms were developed [22-23] to paraphrase rules, where one phrase can be replaced with other having an equivalent meaning. Determining context sensitivity, word sensitivity, incorporating syntax issues [17], style and some of the associated work have been reported in [24-27].

Paraphrases can be generated from images as well. Text description of images can be generated from the features learnt in a deep image recognition model, using either fragment embedding [28] or recurrent neural networks such as long-short term memory (LSTM) ([6],[29]).

One of the most common applications of paraphrasing is query generation and pattern expansion. Query variants are automatically generated and submitted to information retrieval systems or patterns for submission to information extraction systems. Paraphrases have also been used to improve the task of relation extraction [36]. Paraphrases help in answering question and verify the authenticity of answers. Most recently, Bhagat and Ravichandran [31] collected paraphrastic patterns for relation extraction by applying semi-supervised paraphrase induction to a very large monolingual corpus [22]. Another major application area of paraphrasing is summarization of news articles, scientific articles. Mendoza et al. [32] proposed an extractive generic summarization method for single documents by using generic operators and guided local search. This method uses a mimetic algorithm which has combined the population based search of evolutionary algorithm (EA) with a guided local search strategy [33].

There are several open challenges and main problems in this domain such as unavailability of monolingual corpora specific to a domain, expansion of paraphrasing algorithms to several languages, development of integrated model for text-generation tasks such as simplification and sentence compression. Further research in this domain could be in the directions to develop model for low-density languages which have small corpora, to investigate stylistic paraphrasing in other domains, such as paraphrasing emails into

formal documents, or translating legal documents into nontechnical English. Another interesting and major task in this domain is Question generation. Hybrid NMT models using different AI techniques are yet to be developed to generate and answer WH and subjective questions that are syntactically correct.

3.0 Question Generation

Initially question generation has been tackled using rule-based approaches, typically based on deep linguistic knowledge [34-35] such as declarative-to-interrogative sentence transformation. This model was later modified by Heilman and Smith [2] by applying statistical ranking, which ranks multiple questions generated from an input using supervised learning to produce more acceptable questions. Semantic-based parsing models were also built to extract factual statements and modify them to interrogative questions [36-37].

An alternative approach of using generic question templates was developed [38] and the slots are filled with entities from the document. Ontology derived templates [39] were used to generate questions pertaining to larger parts of the document. These approaches contain different tasks that are pipelined but difficult to measure the final outcome.

NN based models that could measure the outcome in terms of rewards were developed to overcome the above drawback. These enable end-to-end training of question generation systems [40]. An encoder-decoder model using a combination of maximum likelihood, policy gradient optimization to generate natural-language questions conditioned on text passages and predefined answers was developed. Reinforce-style algorithm was used to maximize various rewards earned by generated questions. Sequence-to-sequence approach of Cho et al., [41] was adopted for generating questions, conditioned on a document and answer. This model uses human evaluation to determine the correlation between rewards assigned to question and human judgments to improve the model [42] . To overcome the dependency on human judgment, attention mechanisms were introduced to facilitate automatic question generation.

An attention-based sequence learning model investigates the effect of encoding sentence- vs. paragraph-level information [43]. Conditional probability using RNN encoder-decoder architecture [41] was modeled, and global attention mechanism [7] was adopted to make the model focus on certain elements of the input when generating word during decoding. This model encodes sentence-level and paragraph-level information. Using this fully data-driven attention-based neural network approach, automatic question generation was done for reading comprehension.

Another approach for automatic question generation is to extract unigrams, bigrams along with many other components of knowledge base. The sentences are normalized eliminating the sentences that are not eligible for MCQ. Context based similarity approach using paradigmatic relation discovery techniques for generation of options (distracters) other than the correct answer for MCQ. Using this approach, automatic question generation for intelligent tutoring systems was developed for physics domain to generate multiple choice questions given a document.

Question generation systems can also be deployed as chatbot components (e.g., asking questions to start a conversation or to request feedback [44]) or, arguably, as a clinical tool for evaluating or improving mental health [45-46]. Question generation systems can be an integral part in examination systems. Automatic questionnaire can be generated for interviewing candidates based on the qualification and the requirement.

Some of the major challenges in this area are - real time question generation systems given a document in a particular domain, absence of an algorithm or model to verify the authenticity of paraphrased answers, model to ask abstract and interesting questions that are not limited to the document but are within the scope of the document, to deploy attention mechanisms that could understand text and frame question than converting factual statements to interrogative questions. Lately, multi-task learning (MTL) algorithms are used to hybridize the existing algorithms to address the above challenges.

4.0 Multi-Task Learning Algorithms (MTL)

The goal of MTL as summarized by R. Caruana [47] can be stated as: "MTL improves generalization by leveraging the domain-specific information contained in the training signals of related tasks". Multi-task learning has been implemented successfully on various domains of machine learning, natural language processing, drug discovery and computer vision [48-50].

Multi task learning helps the learning model to share representations between related tasks and enable to generalize on the original task. In contrast to single task learning, as soon as we optimize more than one loss function, it implies we are doing multitasking.

MTL is effective due to the below reasons

- Learning a single task has the risk of over fitting but learning two tasks jointly enables the model to obtain better representation by averaging the noise patterns.
- MTL can help the model focus its attention when the task is very noisy or the data is limited or high dimensional.
- MTL facilitates one task to learn from other task.
- It allows generalization of new tasks as MTL biases the model to prefer representations that other tasks prefer as well.

In the context of deep learning, multi-task learning is typically done with either hard or soft parameter sharing of hidden layers [51]. Hard parameter sharing is implemented by keeping task specific output layers and sharing hidden layers between all tasks [52]. This also reduces the risk of over fitting the shared parameters [53]. In contrast to hard parameter sharing, in soft parameter sharing each task has its own model with its own parameters. The distance between parameters of the models is regularized in order to promote the similarity between models. Different distance regularization functions like l_2, trace norm, etc. were reported to be used to promote similarity between models [54-55].

In MTL for computer vision, convolution layers are often shared while learning and task specific are fully connected layers. In addition to this structure, Wang et al. [8] placed matrix priors on fully connected layers. Lu et al., [56] proposed a bottom-up approach for fully-adaptive feature sharing that initializes with a thin network and widens during training using a criterion that groups similar tasks. Misra et al., [57] developed cross-stitch network models. This model contains two separate model architectures and uses cross-stitch units to allow the model to determine the method for task-specific networks and leverage the knowledge of other task by learning combination of output of previous layers.

In contrast, in natural language processing (NLP), recent work focused on finding better task hierarchies for multi-task learning: Søgaard and Goldberg [58], showed that low-level tasks, i.e. NLP tasks typically used for preprocessing such as part-of-speech tagging and named entity recognition should be supervised at lower layers when used as auxiliary task [51]. Based on low supervision model, Hashimoto [59] predefined hierarchical architecture consisting of several NLP tasks and referred it as joint model for multi-task

learning. Hospedales [55] developed matrix factorization approach using tensor factorization to divide the model parameters into shared and task-specific parameters for every layer.

Ruder et al., [60] developed Sluice networks, to learn what layers and subspaces are to be shared, as well as to identify layers that have learned the best representation of the input sequences. This model generalizes Deep Learning based MTL approaches such as hard parameter sharing, cross-stitch networks, block-sparse approaches and NLP approaches that create a task hierarchy.

MTL was implemented in several scenarios in the domain of machine comprehension:
A multi-task DNN approach was developed by Liu. X. et al. [61], for query classification and information retrieval. It maps arbitrary text queries and documents into semantic vector representations in a low dimensional latent space. The model combines ranking tasks and classification. The hidden layers are shared, task-specific representation is implemented and the model is trained using mini-batch-based stochastic gradient descent algorithm.

Multi-tasking DNN was developed to identify a valid comment for new questions from the existing comments for old questions in community question answering (cQA) portal. This task can be divided into three subtasks.
(A) Predict if a comment produced in response to a question contains a valid answer;
(B) Re-rank a set of questions according to their relevance with respect to the original question;
(C)Predict if a comment produced in response to a previous question posted on the cQA forum represents a valid answer to a fresh question.
MTL approach produces higher accuracy and faster convergence rates than individual neural networks. The neural architecture consists of CNN sentence encoder which is a convolution operation followed by a k max pooling layer and provides comparable performance to LSTM on the task of new question-comment [62].

Multi-column convolutional neural networks (MCCNNs) were introduced to understand questions from three different aspects (answer path, answer context, and answer type) and learn their distributed representations. Question paraphrases were leveraged to train the column networks in a multi-task learning manner. Learning is done using low-dimensional embeddings for the entities and relations in FREEBASE. Question-answer pairs were provided as supervision signals. An intuitive interpretation for MCCNNs were provided by conducting extensive experiments on WEBQUESTIONS dataset and developing method to detect important question words in the different column networks [55].

Given a natural language question, Li et al., [63] developed a model that estimates a probability distribution over candidate answers. Questions are paraphrased using one or several paraphrasing systems and neural scoring model predicts the quality of the generated paraphrases. QA model also predicts a distribution over answers given the question and its paraphrased versions. The system is trained using question-answer pairs as a supervision signal and is independent of any specific paraphrase or QA models. The framework can be used for other natural language processing tasks which are sensitive to the variation of input (e.g., textual entailment or summarization). More advanced paraphrase scoring models [64-65] as well as additional paraphrase generators could improve the diversity and quality of paraphrase and enhance QA performance. As per the literature review, MTL had been applied only for query classification and identification. We intend to apply multi-task learning to develop a hybrid model for paraphrasing and query generation.

5.0 Identified Challenges

A review of various text processing applications like question generation, paraphrasing algorithms and multi-task learning algorithms has been discussed in the text above. The review also identified the limitations and challenges of the aforementioned approaches; it therefore opens up new avenues of research which is discussed below.

One of the major applications of paraphrasing algorithms is question generation. In the current scenario, questions are generated by extraction of factual statements and application of statistical or AI techniques to paraphrase them into interrogative sentences. This leads to the first objective of the proposal; to study the existing paraphrasing and question generation algorithms. The study aims to identify the drawbacks of existing algorithms and provide inputs to design hybrid technique.

Paraphrasing is implemented using statistical techniques like phrase-based SMT, syntax-based SMT and soft computing techniques like NMT, EA with guided local search, etc. The second objective aims to explore the possibility of hybridizing existing algorithms using advanced techniques like Quantum EA, Reinforcement learning techniques, etc, to address open challenges such as question generation, text summarization and stylistic paraphrasing.

Based on the inputs from the first two objectives, we intend to address the major challenges in question generation models. Current models lack attention mechanisms that can understand text and frame syntactically correct questions. They lack the skill to frame context sensitive subjective questions. The last objective aims to explore the possibility of implementing hybrid model to address the above challenges. Such models can also be used in automation of recruitment system, examination system, personality assessment system, speech therapy systems, etc. As a case study, we will explore the possibility of implementing the model for automatic question generation from audio file. In our university (Dayalbagh Educational Institute), daily home assignments (DHA) which consist of few questions are given to the students as assignment after every lecture for continuous learning and assessment. We attempt to implement our hybrid model with required changes to generate questions from the lecture delivered in the class in real-time.

This leads to summarize our research proposal to the following objectives.

6.0 Proposed Research Objectives

1. Investigation of multiple paraphrasing algorithms and their effectiveness.
2. To explore the possibility of designing a hybrid soft computing model for paraphrasing using different techniques.
3. Comparative analysis and implementation of above hybrid model for applications like question generation, summarization to address context sensitivity and syntax challenges.

It is proposed to limit our research in this thesis to an appropriate mixture and subset of the issues identified above.

References

[1] W. Xu et al., “Optimizing Statistical Machine Translation for Text Simplification”, Transactions of the Association for Computational Linguistics, Vol. 4, pp. 401-415, 2016.
[2] M. Heilman and N. A. Smith, “Good Question! Statistical Ranking for Question Generation”, HLT ‘10

Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 609-617, 2010.
[3] J. Mallinson et al., “Paraphrasing Revisited with Neural Machine Translation”, In Proc. 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pp. 881–893, Valencia, Spain, 2017.
[4] N. Kalchbrenner and P. Blunsom, “Recurrent continuous translation models”, In Proc. of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1700–1709, Seattle, Washington, USA, 2013.
[5] D. Bahdanau et al., “Neural machine translation by jointly learning to align and translate”, arXiv:1409.0473v7, 2016.
[6] I. Sutskever et al., “Sequence to sequence learning with neural networks”, In Advances in Neural Information Processing Systems, Vol. 27, pp. 3104–3112. Curran Associates, Inc., 2014.
[7] T. Luong et al., “Effective approaches to attention-based neural machine translation”, In Proc. Conference on Empirical Methods in Natural Language Processing, pp. 1412–1421, Lisbon, Portugal, 2015.
[8] T. Wang et al., “A Joint Model for Question Answering and Question Generation”, arXiv preprint arXiv:1706.01450, 2017.
[9] J. Ganitkevitch et al., “Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation”, In Proc. Conference on Empirical Methods in Natural Language Processing, pp. 1168–1179, Edinburgh, Scotland, UK, 2011.
[10] A. Ibrahim et al., “Extracting structural paraphrases from aligned monolingual corpora”, In Proc. Second International Workshop on Paraphrasing, Sapporo, Japan, 2003.
[11] L. Iordanskaja et al., “Lexical selection and paraphrase in a meaning text generation model”, in Natural Language Generation in Artificial Intelligence and Computational Linguistics, Kluwer Academic, Chapter 11, 1991.
[12] K. R. McKeown et al., “Tracking and summarizing news on a daily basis with Columbia’s Newsblaster”, In Proc. Human Language Technology Conference, San Diego, 2002.
[13] B. Pang et al., “Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences”, In Proc. Human Language Technology Conference, 2003.
[14] R. Barzilay and K.McKeown, “Extracting paraphrases from a parallel corpus”, In Proc. Association for Computational Linguistics, pp. 50-57, 2001.
[15] R. Barzilay and L.Lee, “Learning to paraphrase: An unsupervised approach using multiple sequence alignment”, In Proc. Human Language Technology Conference, 2003.
[16] P. Koehn et al., “Statistical phrase-based translation”, In Proc. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Vol. 1, pp. 48–54, Sapporo, Japan, 2003.
[17] C. Callison, “Syntactic constraints on paraphrases extracted from parallel corpora”, In Proc. Empirical Methods in Natural Language Processing, pp. 196– 205, Honolulu, Hawaii, 2008.
[18] J. Mallinson et al., “Paraphrasing Revisited with Neural Machine Translation”, In Proc. Conference of the European Chapter of the Association for Computational Linguistics: Vol. 1, Long Papers, pp. 881–893, Valencia, Spain, April 3-7, 2017.
[19] J.Ganitkevitch et al., “The paraphrase database”, In Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 758–764, Atlanta, Georgia, 2013.
[20] J. Ganitkevitch et al., “The multilingual paraphrase database”, In the 9th edition of the Language Resources and Evaluation Conference, Reykjavik, Iceland, 2014.

[21] X. Zhang and Y. LeCun, “Text Understanding from Scratch”, arXiv:1502.01710v5, 2016.
[22] N. Madnani and B.J. Dorr, “Generating phrasal and sentential paraphrases: A survey of data- driven methods”, Computational Linguistics, Vol. 36 (3), pp. 341-387, 2010.
[23] I. Androutsopoulos and P. Malakasiotis, “A survey of paraphrasing and textual entailment methods”, Journal of Artificial Intelligence Research, pp. 135–187, 2010.
[24] R. Bhagat and E. Hovy, “What is a paraphrase?” , Computational Linguistics, Vol. 39 (3), pp. 463–472, 2013.
[25] M. Apidianaki, E. Verzeni, and D. Mc-Carthy, “Semantic clustering of pivot paraphrases”, International Conference on Language Resource and Evaluation, 2014.
[26] E. Pavlick and A. Nenkova, “Inducing lexical style properties for paraphrase and genre differentiation”, In North American Chapter of the Association of Computational Linguistics (NAACL): Human Language Technologies, 2015.
[27] W. Xu et al., “Paraphrasing for style”, In Computational Linguistics (COLING), pp. 2899–2914, 2012.
[28] A. Karpathy and F. Li, “Deep visual-semantic alignments for generating image descriptions”, arXiv:1412.2306, December 2014.
[29] K. Xu et al., “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention”, In Proc. 32nd International Conference on Machine Learning, Lille, France, 2015.
[30] Romano et al., “Investigating a generic paraphrase-based approach for relation extraction”, In Proc. EACL, pp. 409–416, Trento, 2006.
[31] R. Bhagat and D. Ravichandran, “Large scale acquisition of paraphrases for learning surface patterns”, In Proc. ACL, pp. 674–682, Columbus, OH, 2008.
[32] M. Gambhir and V. Gupta, “Recent automatic text summarization techniques: a survey”, Artificial Intelligence Review, Vol. 47(1), pp. 1-66, 2017.
[33] M. Mendoza et al., “Extractive single-document summarization based on genetic operators and guided local search”, Expert Systems with Applications, Vol. 47, pp. 4158-4169, 2014.
[34] R. Mitkov and L. A. Ha., “Computer- aided generation of multiple-choice tests”, In Proc. HLT-NAACL 03 Workshop on Building Educational Applications Using Natural Language Processing, pp. 17–22, 2003.
[35] V. Rus et al., “The first question generation shared task evaluation challenge”, In Proceedings of the 6th International Natural Language Generation Conference. ACL, Stroudsburg, PA, USA, pp. 251–257, 2010.
[36] P. Mannem, R. Prasad, and A. Joshi, “Question generation from paragraphs at UPenn: QGSTEC system description”, In Proceedings of QG2010: The Third Workshop on Question Generation, pp. 84–91, 2010.
[37] D. Lindberg et al., “Generating natural language questions to support learning on-line”, Proc. of ENLG, 2013.
[38] Y. Chali and S. Golestanirad, “Ranking automatically generated questions using common human queries”, Proc. of INLG, 2016.
[39] I. Labutov, S. Basu, and L. Vanderwende, “Deep questions without deep understanding”, Proc. of ACL, 2015.
[40] I. V. Serban et al., “Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus”, Proc. of ACL, 2016.
[41] K. Cho et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation”, In Empirical Methods on Natural Language Processing, 2014.
[42] X. Yuan, et al., “Machine Comprehension by Text-to-Text Neural Question Generation”, arXiv:1705.02012v2, 2017.

[43] Du. X., et al., “Learning to Ask: Neural Question Generation for Reading Comprehension”, arXiv:1705.00106v1, 2017.
[44] Mostafazadeh et al., “Generating natural questions about an image”, arXiv preprint arXiv:1603.06059, 2016.
[45] J. Weizenbaum, Eliza & mdash, “A computer program for the study of natural language communication between man and machine”, Commun. ACM, Vol. 9 (1), pp. 36–45, 1966.
[46] K. M. Colby, S. Weber, and F. D. Hilf, “Artificial paranoia”, Artificial Intelligence, Vol. 2 (1), pp. 1–25, 1971.
[47] R. Caruana, “Multitask Learning. Autonomous Agents and Multi-Agent Systems”, Vol. 27 (1), pp. 95–133, 1998.
[48] R. Collobert et al., “Unified architecture for natural language processing”, Proc. International conference on Machine learning - ICML ’08, Vol. 20 (1), pp. 160–167, 2008.
[49] B. Ramsundar et al., “Massively Multitask Networks for Drug Discovery”, arXiv:1502.02072v1, 2015.
[50] R. Girshick, “Fast R-CNN”, In Proc. IEEE International Conference on Computer Vision, pp. 1440–1448, 2015.
[51] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks", arXiv:1706.05098v1, 2017.
[52] Caruana, "Multitask learning: A knowledge-based source of inductive bias”, In Proc. Tenth International Conference on Machine Learning, 1993.
[53] J. Baxter, " A Bayesian/information theoretic model of learning to learn via multiple task sampling”, Machine Learning, Vol. 28, pp.7–39, 1997.
[54] Duong et al., "Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser”, In Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers), pp.845–850, 2015.
[55] Y. Yang and T. M. Hospedales, “Trace Norm Regularised Deep Multi-Task Learning”, In Workshop track - ICLR, 2017.
[56] Lu, Y., Kumar et al., “Fully- adaptive Feature Sharing in Multi-Task Networks with Applications in Person Attribute Classification,” In Proc. Conference on Computer Vision and Pattern Recognition, 2016.
[57] I. Misra et al., “Cross-stitch Networks for Multi-task Learning”, In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[58] A. Søgaard and Y. Goldberg, “ Deep multi-task learning with low level tasks supervised at lower layers”, In Proc. 54th Annual Meeting of the Association for Computational Linguistics, pp. 231–235, 2016.
[59] K. Hashimoto, "A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks”, arXiv:1611.01587v5, 2017.
[60] S.Ruder et al., “Sluice networks: Learning what to share between loosely related tasks”, arXiv:1705.08142v2, 2017.
[61] X Liu et al., “Representation learning using multi-task deep neural networks for semantic classification and information retrieval”, In Proc. NAACL, 2015.
[62] D. Bonadiman, "Multitask Learning with Deep Neural Networks for Community Question Answering",arXiv:1702.03706v1, 2017.
[63] L. Dong, "Learning to Paraphrase for Question Answering”, arXiv:1708.06022v1, 2017
[64] A. Parikh et al., “A decomposable attention model for natural language inference”, In Proc. Empirical Methods in Natural Language Processing, pp. 2249–2255, Austin, Texas, 2016.

[65] S. Wang and J.Jiang, “Learning natural language inference with lstm”, In Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1442– 1451, San Diego, California. ACL, 2016.